How bacterial ribosomal protein L20 assembles with 23 S ribosomal RNA and its own messenger RNA.

Sophie Raibaud, Patrice Vachette, Maude Guillier, Frédéric Allemand, Claude Chiaruttini, Frédéric Dardel

How bacterial ribosomal protein L20 assembles with 23S ribosomal RNA and its own messenger RNA.

Sophie RAIBAUD[1], Patrice VACHETTE[2], Maude GUILLIER[3], Frédéric ALLEMAND[3], Claude CHIARUTTINI[3] & Frédéric DARDEL[1]*

[1]Cristallographie & RMN Biologiques, UMR 8015 CNRS, Faculté de Pharmacie, 4 avenue de l'Observatoire, 75006 Paris, France. [2]LURE, UMR 00130 CNRS, Université Paris-sud, B.P. 34, 91898 Orsay Cedex, France. [3]Institut de Biologie Physico-Chimique, UPR9073 CNRS, 13 rue Pierre et Marie Curie, 75005 Paris, France.

Running title: L20-RNA interactions

* Corresponding author:

E-mail: dardel@pharmacie.univ-paris5.fr

Telephone: +33 1 53 73 99 93

Fax: +33 1 53 73 99 25

Keywords: Ribosome, translational control, L20, NMR, SAXS

Document Word Count: 8501

Summary.

In bacteria, the expression of ribosomal proteins is often feedback regulated at the translational level by the binding of the protein to its own mRNA. This is the case for L20, which binds to two distinct sites of its mRNA that both resemble its binding site on 23S rRNA. In the present work, we report an NMR analysis of the interaction between the C-terminal domain of L20 (L20C) and both its rRNA and mRNA binding sites. Changes in the NMR chemical shifts of the L20C backbone nuclei were used to show that the same set of residues are modified upon addition of either the rRNA or the mRNA fragments, suggesting a mimicry at the atomic level. In addition, SAXS experiments, performed with the rRNA fragment, demonstrated the formation of a complex made of two RNAs and two L20C molecules. A low-resolution model of this complex was then calculated using (i) the rRNA/L20C structure in the 50S context and (ii) NMR and SAXS results. The formation of this complex is interesting in the context of gene regulation since it suggests that translational repression could be performed by a complex of two proteins each interacting with the two distinct L20 binding sites within the operator.

Introduction.

Ribosome assembly is a complex process in which proteins must assemble onto the ribosomal RNA in an ordered fashion. This process has been extensively analysed for bacterial ribosomal subunits, using *in vitro* reconstitution studies. In the case of *Escherichia coli* large 50S subunit, it has been shown that nine "core proteins" L1, L3, L4, L9, L10, L11, L20, L23 and L24 bind directly to the 23S rRNA, independently of other proteins (1,2). The other 50S ribosomal proteins (r-proteins) depend upon those primary binders for attachment to the large ribosomal subunit. Interestingly, several of these core proteins, namely L1, L4, L10 and L20, are also involved in the feedback regulation that allows the coordinated expression of the various ribosomal components (3-8). A similar situation is also observed for the small ribosomal subunit where regulatory r-proteins are also primary 16S rRNA binders.

Most bacterial ribosomal protein genes are clustered in polycistronic operons, the expression of which is controlled at the translational level(9-11). One cistron encodes a regulatory ribosomal protein that binds to the messenger RNA, thereby shutting off the translation of the downstream r-proteins cistrons. In the classical Nomura model, it is assumed that the protein binding site on the mRNA, the translational operator, is structurally similar to its binding site on the rRNA. Molecular mimicry between the two target RNA sites would then be used to adjust the translation of r-proteins to the level of transcription of the rRNA: in the presence of excess unbound rRNA, the repressor r-protein would be displaced from its mRNA, thereby allowing translation to proceed. So far, in *E. coli*, the molecular mimicry between the translational operator on the mRNA and the rRNA has been demonstrated in the case of L1, through mutagenesis of either target sites (12), and of S8 where direct binding to both RNAs has additionally been shown(13).

More recently, detailed studies were performed on the feedback control mechanism of S15 from various bacterial species. In *E. coli*, S15 translational operator is a pseudoknot structure (14), whereas in *T. thermophilus* and *B. stearothermophilus* it is a three-helix junction, as in the binding site for S15 in central domain of the 16S rRNA (15,16). However, in all three cases, there appear to be a common recognition pattern: a bipartite recognition motif with a conserved G-U/G-C in one of the two subsites (15-17). This motif is also involved in S15 binding to the16S rRNA, as seen in the crystal structures (18,19), thereby confirming the molecular mimicry model.

In *E. coli*, the case of L20, one of the four 50S repressor proteins, is quite puzzling because of the unique nature of its translational operator organisation. The L20 gene is part of the IF3 operon, which comprises *infC*, *rpmI* and *rplT*, encoding translation initiation factor IF3, and ribosomal proteins L35 and L20, respectively. L20 represses the translation of *rpmI* (8), which in turn prevents the translation of its own gene through translational coupling (20). Regions of the *rpmI* leader sequence important for translational control have been identified by a combination of genetic, mutational and footprinting experiments. They are composed of two independent structures: a pseudoknot structure which involves long-range RNA-RNA interactions and overlaps the *rpmI* ribosome binding site and initiation codon (site 1) (21), and an imperfect stem structure located in the *infC-rpmI* intergenic region (site 2) (22). Interestingly, both sites are apparently independently bound by L20, and both are required for repression of expression of L35 and L20 (22). This peculiar “dual-site” organisation of the operator raises a number of questions: (i) How many L20 molecules are required to bind to each site? (ii) Why is simultaneous binding required for repression? (iii) What is the mechanism allowing the crosstalk between the two sites?

The 3D structures of free and ribosome-bound forms of L20 have been recently reported (23,24). They reveal a segmented organisation, with a globular C-terminal domain (L20C) which interacts with helices 40 and 41 of the 23S rRNA (25), and a highly cationic N-terminal domain, which is disordered in the free state and folds as a long helical shaft within the ribosome, penetrating deeply into the 50S subunit(24). Interestingly, it has been shown that, by itself, L20C from *E. coli* is able to act as a translational repressor *in vivo*(23). Therefore, the N-terminal domain of L20 appears to be dispensable for the control of *rpmI* and *rplT* expression. The present work reports the study of the interaction of L20C with either its target site on the 23S rRNA or one of its two binding sites within the translational operator of *rpmI*, using heteronuclear NMR and SAXS[1]. Both selected RNA fragments interact specifically with L20C and the two complexes involve similar residues within the protein, indicating a common binding site. The observed complexes are however significantly larger than expected and the SAXS data suggest that it is a dimer of complexes *i.e.* 2 proteins+2 RNAs. This indicates that L20C has the intrinsic capacity to dimerise RNA and sheds new light on the translational control mechanism. In the case of the rRNA/protein complex, a low-resolution structural model of this dimer could be constructed from both the NMR and SAXS data, yielding some structural indications on how L20 could interact with both sites on its messenger RNA.

[1] Abbreviations : EDTA, Ethylene diamine tetraacetic acid; HSQC, Heteronuclear single quantum correlation; NMR, Nuclear magnetic resonance; NOESY, Nuclear Overhauser effect spectroscopy; PAGE, Polyacrylamide gel electrophoresis; PMSF, Phenylmethanesulphonyl fluoride; SAXS, Small angle X-ray scattering; TOCSY, Total correlation spectroscopy; TROSY, Transverse relaxation optimised spectroscopy; SDS, Sodium dodecylsulphate.

Experimental procedures

Protein expression and purification

The plasmid pET42aL20aaΔN was constructed as follow: pBL20aaΔN (M. Guillier *et al.*, manuscript in preparation) was digested by *Bsr*GI-*Xba*I and the DNA fragment carrying L20C was transferred into pET42a (Novagen). The resulting plasmid was transformed into BL21 (DE3) CodonPlus (Stratagene).

For expression and labelling of L20C, BL21 (DE3) CodonPlus/pET42aL20aaΔN was grown at 37°C in Martek-9 N medium (Spectra Stable Isotopes) supplemented with Celtone-N (2g/l), and the appropriate antibiotics. When the A_{600} turbidity reached 0.5, L20C expression was induced by addition of 0.5 mM IPTG and protein production was allowed to proceed for 4 hours. Typical cultures (0.5 litre) yielded approximately 2 g of cells (wet weight). Cells were harvested by centrifugation, suspended in 10 ml of 200 mM NaCl, 50 mM Tris-HCl pH 7.5, 1 mM EDTA, 1 mM PMSF, and disrupted by sonication. Cell debris were spun down and the soluble extract was concentrated by ammonium sulphate precipitation (70% saturated). The sample was first submitted to a gel filtration column (Superdex75 prep-grade 2.6 x 60 cm, Amersham) equilibrated in 20 mM Tris-HCl, 500 mM NaCl, 0.5 mM EDTA, pH 7.5, at room temperature. The presence of L20C within the collected fractions (4ml each) was monitored by analysing aliquots on SDS-PAGE. The pooled fractions (24 ml) were then diluted 2.5 times with H_2O to lower the ionic strength. An ion-exchange chromatography step was then performed on an SP Sepharose column (Hiload 2.6 x 10 cm, Amersham) equilibrated in 20 mM Tris-HCl, 200 mM NaCl, pH 7.5. L20C was eluted using a 0.2-1.5 M linear NaCl gradient in the same buffer (300 ml). The purified protein was then dialysed against 20 mM potassium phosphate, 100 mM NaCl,

pH 6.6, and concentrated with a centricon YM-3 (Millipore) The overall yield was 8 mg of purified L20C per litre of culture.

RNA fragments corresponding to the rRNA and oRNA target sites (Fig. 1) were purchased from Dharmacon Research, deprotected as indicated by the supplier and dialysed extensively. $tRNA^{Lys}_3$ was purified as previously described (26).

NMR titration

The NMR experiments were carried out on a Bruker DRX600 Avance spectrometer. NMR titrations were performed using a 420 µl sample at a final L20C concentration of 0.6 mM in buffer 20 mM potassium phosphate, 100 mM NaCl, 0.1 mM EDTA, pH 6.6. One molar equivalent of RNA was lyophilised and suspended in 80µl of the same buffer. Aliquots of this solution were added stepwise to the labelled L20C to perform a several point titration. RNA/L20C molar ratios used were 0.5:1, 0.6:1, 0.8:1 and 1:1. The pH was checked after each step and remained within 0.1-0.2 pH unit of the starting value. NMR spectral changes were monitored by recording TROSY experiments (27) at 303 K. The assignment of the L20C residues in the complex was derived from a 3D-NOESY-TROSY experiment (28) with a mixing time of 80 ms, also at 303 K. NMR data were processed with the Gifa software (29).

Gel filtration and SAXS experiments

The complex was formed as described above, by stepwise addition of the RNA to the protein solution, and monitored by NMR. The complex recovered from the NMR tube (6.5 mg in 0.5 ml) was submitted to a gel filtration step (Sephacryl S-100 HighPrep 2.6 x 60 cm, Amersham) running at 0.1 cm/min in 20 mM potassium phosphate, 100 mM NaCl, 0.1 mM EDTA, pH 6.5, at 4°C. Fractions (3.6 ml) were analysed by SDS-PAGE and urea-PAGE for protein and RNA

content, respectively. The combined complex fractions (36 ml) were then concentrated to 5 mg/ml by ultrafiltration using ultrafree-5K cells (Millipore).

SAXS measurements were performed using the synchrotron radiation beam line D24 at the DCI storage ring of LURE (Laboratoire pour l'Utilisation du Rayonnement Électromagnétique, Orsay, France). The instrument, the data acquisition system (30) and the evacuated measuring cell (31) have been described. Measurements were performed in a quartz capillary (diameter *ca* 1.5 mm) maintained at the constant temperature of 20°C on solutions of free rRNA and L20C/RNA complex in 20 mM potassium phosphate, 100 mM NaCl, 0.1 mM EDTA, pH 6.5. Samples were at a macromolecular concentration close to 5 mg/ml. Eight frames of 100 s each were recorded using a position sensitive proportional detector covering the range of modulus of the scattering vector s from 0.0016 to 0.03 $Å^{-1}$ ($s = 2 \sin\theta/\lambda$, where 2θ is the scattering angle, λ is the radiation wavelength, λ=1.488 Å). Eight frames of 200 s each were also recorded with the corresponding buffer. Frames were visually inspected to check for X-ray damage; none was found. Data were scaled to the transmitted intensity, before computing the average and standard deviation of each measurement and subtracting the scattering from the corresponding buffer. The scattering pattern of a 5 mg/ml solution of BSA was also recorded and the value of the intensity at the origin I(0)/c (c protein concentration in mg/ml) was used as a reference to derive an estimate of the molecular mass of the samples studied.

The structural parameters were calculated using the following standard procedures. The radius of gyration and the intensity at the origin were derived from a Guinier analysis of the scattering intensities at $s < s_{max} = x/(2\pi R_g)$ (32) (x=1.3 for the RNA, x=1.55 for the complex) :

$$\ln[I(s)] = \ln[I(0)] - (4\pi^2 R_g^2/3)\, s^2$$

The distance distribution function *p(r)* corresponds to the distribution of distances between any two volume elements within one particle. It has been determined using the indirect transform method as implemented in the program GNOM (33).This function provides an alternative estimate of the radius of gyration derived through the relationship:

$$R_g^2 = \frac{\int r^2 p(r) dr}{2\int p(r) dr}$$

Scattering intensities were computed from the atomic coordinates of L20C, rRNA and their complexes by using the program CRYSOL, which takes into account the hydration water by introducing a 3 Å thick border layer surrounding the molecule (34). The calculated scattering profile is fitted to the experimental pattern using only two adjustable parameters, the average displaced solvent volume per atomic group and the contrast of electron density of the border layer with respect to bulk solvent $\Delta\rho_b = \rho_b - \rho_0$, to minimize the discrepancy:

$$\chi^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(\frac{I_e(s_i) - I(s_i)}{\sigma(s_i)}\right)^2$$

where N is the number of experimental points, and $I_e(s_i)$ and $\sigma(s_i)$ denote the experimental scattering curve and its standard deviation, respectively.

Molecular modelling

In the following text, "complex" refers to the L20C/rRNA fragment complex (1 protein+ 1 RNA duplex), whereas "dimer" designates the dimer of complexes (2 proteins + 2 RNA duplexes). The structure of the complex between L20C and the rRNA fragment was modelled as follows: coordinates from L20C and the rRNA fragment were extracted from *D. radiodurans* 50S ribosomal subunit structure (24) (the latest PDB entry,1NKW, was used, as the original 1KC9

entry contained a 3 residue shift in the chain tracing within L20). Since only the Cα atoms of the protein were included in this structure, the complete coordinates from L20C from *A. aeolicus* (23) (PDB 1GYZ) were then substituted for the *D. radiodurans* protein, after least-square fitting Cα atom positions of the four α-helices composing the fold. Non-polar hydrogen atoms were removed and the docked complex was energy-minimized in XPLOR(35) using the CHARMM force field(36) in order to remove bad van der Waals contacts.

The sampling of allowable dimer conformations was performed using XPLOR. (i) The L20C/RNA complex structure was translated so that its centre of mass coincided with the origin of the Cartesian coordinates. (ii) A random axis going through the origin was generated. The L20C/RNA complex structure was duplicated in place and the second complex was rotated by 180 degrees about this random axis. (iii) A random unit vector, perpendicular to the rotation axis, was then generated and the second complex was translated by 100 Å along this axis, a distance sufficient to ensure that both dimers were no longer in contact with each other. (iv) The second complex was then moved back stepwise along the translation axis, until it came back in contact with the first complex. Contact between the two complexes was detected by monitoring the van der Waals term of XPLOR energy function. After these steps, a random dimer conformation was obtained in which the two contacting complexes were related by a two-fold symmetry axis.

For each random dimer thus generated, the following three geometrical parameters were computed: the radius of gyration, the minimum distance between L20C in the first complex and the RNA in the second complex (the *bridging distance*), and the minimum distance between the C-terminal helix of L20C (residues 102-118) in the first complex and the closest atom in the second complex (either from L20C or from the RNA (the *C-terminus distance*). When the resulting dimer geometry met the constraints, ***i.e.*** the three geometrical parameters fell within

preset intervals, the coordinates were written to a file, otherwise the dimer was rejected. The set of randomly generated dimers that fitted the imposed geometrical constraints (compliant dimers) was then clustered into families, based on their r.m.s.d computed over all atoms. Any two dimer structures that exhibited a global pairwise r.m.s.d under 6 Å were considered neighbours. This neighbouring relationship could be used to partition the set of compliant dimers into disjoint geometric clusters.

Results.

Structural analysis of L20C, the translational control domain.

In order to investigate the recognition of *rpmI* translational operator, we decided to focus on the C-terminal domain of the protein, L20C, which corresponds to the second half of the sequence, starting approximately at the single tryptophan residue in the middle of the protein. Interaction studies were thus not conducted with the full length L20 for the following reasons: (i) The C-terminal domain suffices to repress *rpmI* expression (23) (ii) In the crystal structure of *Deinococcus radiodurans* 50S subunit(24), the C-terminal domain of L20 interacts tightly with a discrete region of the 23S rRNA, at the junction between helices 40 and 41(25), which exhibits similarities with the *rpmI* translational operator(22). (iii) It was suspected that the strongly basic and unfolded N-terminal domain would make spurious electrostatic contacts with the RNA fragments, yielding to non-specific aggregation of the sample, a phenomenon indeed observed in preliminary experiments (not shown).

Preliminary studies were conducted with the L20C domain from *E. coli*, in order to study the cognate systems: *E. coli* L20/*E. coli* rRNA and mRNA. However, both the full length L20 and the L20C domain from *E. coli* appeared unstable and gave poor quality NMR spectra (not shown). As we previously solved the solution structure of the *Aquifex aeolicus* L20 protein (L20aa), which is highly homologous to its *E. coli* counterpart (~66 % sequence identity) it was decided to use the L20aa protein to study the interaction with *E. coli* 23S rRNA and *rpmI* mRNA. Furthermore, the expression of *E. coli rpmI* gene can be regulated *in trans* by L20aa (M. Guillier, to be published). Such a heterologous regulation ability had already been demonstrated in *E. coli* for L4 proteins from various bacterial species (37). Finally, L20aa is able to complement a null

mutant of L20 from *E. coli* (M. Guillier, to be published). Thus, L20aa is a suitable model for structural studies, as it has been shown to be able to substitute for both *E. coli* L20 functions *in vivo*: translational control and ribosome assembly.

The L20C domain from *A. aeolicus* was expressed from a recombinant construct: the first 5 codons of the gene were retained and fused to the C-terminal domain, two residues upstream from the tryptophan, yielding the following N-terminal sequence: MRVKGRLW, where the underlined residues correspond to the start of L20C sequence. The protein, which is 67-residue long, was expressed and purified as described under experimental procedures.
In a preliminary NMR study, 3D TOCSY [^{15}N,^{1}H] HSQC and 3D NOESY [^{15}N,^{1}H] HSQC spectra were recorded on the ^{15}N-labelled domain. Interestingly, even in the absence of its N-terminal part, the L20C domain folding remains intact, as evidenced by an almost identical [^{15}N,^{1}H] HSQC pattern, and very similar NOE cross-peaks, although the lines were significantly sharper, as expected for a smaller protein. The complete chain tracing from W59 to Q118 was verified and allowed the identification of a few additional amide groups (Fig. 2A, W59, I60, E87 and R90, full-length L20aa numbering), which were broadened to baseline in the full-length protein. Therefore, the NMR structure and the complete sequence-specific NMR assignments of L20C provided a basis for structural studies of the interactions between L20C and its RNA substrates.

Target RNA sequences

Two RNA fragments were designed according to the L20C-binding sites, either on the 23S rRNA (rRNA) or site 2 on the *rpmI* translational operator (22) (oRNA). Both sites, which were

previously reported to be structurally similar, were constructed by assembling two partially complementary synthetic oligoribonucleotides (Fig 1). The operator sequence (oRNA) top helix was further stabilised by two additional GC base pairs. After annealing of their strands, the two RNA fragments (rRNA: 42 nucleotides, oRNA: 41 nucleotides) where analysed by NMR in the absence of magnesium ions. The imino region of the 1D proton spectra of the rRNA and the oRNA display 10 and 11 resonances respectively (not shown), in good agreement with the predicted RNA secondary structures. In addition, the 2D NOESY spectrum of the oRNA reveals a strong cross peak between two imino protons characteristic of a U:G wobble pair, as expected from the predicted interaction between G21 and U30 (Fig. 1, oRNA). Most of the base and anomeric protons of the two rRNA stems were assigned using 2D NOESY in H_2O and ^2H_2O. The NOE connectivities between base H8/H6 and ribose H1' indicate classical A-form RNA stems as expected from the secondary structures (not shown). Thus both RNAs fold according to the predicted secondary structure, even in the absence of magnesium. All further experiments were performed at 303K, as it appeared to be the highest temperature at which all imino resonances were observable.

L20C binds both RNA via the same site

Titration of ^{15}N labelled L20C with non-labelled RNA was followed with [^{15}N,^1H] correlation spectroscopy. L20C solutions were titrated by stepwise addition of RNA, as described in experimental procedures, and spectral changes were followed using TROSY experiments (figures 2B,C, D). The binding behaviour of both rRNA and oRNA to L20C is qualitatively very similar. First, upon addition of either the rRNA or the oRNA fragment, the protein signals disappear until an RNA/protein ratio of 0.5 is reached, presumably because of unspecific electrostatic binding of excess basic protein onto the acidic RNAs. Indeed, the RNA fragments are about twice the size of

the L20C protein (7.6 kDa for the protein, ~13.5 kDa for the RNAs) and their elongated shape should offer numerous non-specific sites in addition to the specific one. However, when the RNA/protein ration reached 0.5 and was then further increased, the signals reappeared and dramatic changes were observed in the TROSY spectra (Fig. 2). New cross-peaks appeared at the expense of those assigned to the free protein, indicating that, in both cases, binding onto the specific site occurred in the slow exchange regime on the chemical shift time scale, with an estimated lower limit of 60 ms for the half-life of the complex. At an rRNA/L20C ratio equal to one, all cross-peaks corresponding to L20C completely disappeared, while those corresponding to rRNA/L20C reached a maximum intensity. For the oRNA/L20C, disappearance of the free L20C peaks was however not complete. However, this did not impair the analysis of the data, as the bound form of L20C is in slow exchange on the NMR timescale with the free form and the two independent sets of peaks can be easily distinguished. The fact that the oRNA/L20C spectrum still contains free and bound L20C signals (see E87 in Fig. 2D) is probably the consequence of some heterogeneity of the oRNA structure. It was indeed observed that the "long" strand of the oRNA (nucleotides 1-25) can self-pair (F. Allemand, unpublished) and a small amount (~10-20 %) of misfolded RNA is always present in annealed oRNA samples.

The spectral changes observed in the rRNA/L20C and oRNA/L20C complexes appear strikingly similar. This is illustrated in Fig. 2, which shows a side-by-side view of the [^{15}N,^{1}H] TROSY spectra of free L20C (Fig. 2A) and L20C in complex with rRNA (Fig. 2C) or oRNA (Fig. 2D). Very large changes are observed in the ^{15}N and ^{1}H backbone chemical shifts throughout L20C, for both rRNA and oRNA binding. Peaks are affected upon rRNA or oRNA addition in a similar way : (i) L20C amide cross-peaks that are only marginally affected upon RNA binding are exactly the same in both complexes (see G85, V98, A103, F104, E105, Q 106, L115 in fig 2C

and 2D). (ii) Some cross-peaks completely disappear upon either rRNA or oRNA addition (see for instance V67, G71, Y74, S75, and T76 in fig 2C and 2D), while new, strongly shifted peaks appear at similar positions in the spectra of both complexes (boxed in fig. 2C and 2D). These results provide evidence for a direct interaction between both RNA and L20C and further suggest that the chemical environment of the L20C residues is very similar in both rRNA/L20C and oRNA/L20C complexes. As a negative control, the interaction of L20C with a non-cognate RNA, human $tRNA^{Lys}_3$, was also monitored by a similar procedure. In contrast to what was observed with oRNA and rRNA, very little change is observed in the TROSY spectrum upon addition of the tRNA (fig 2B). Superimposition of the [^{15}N,^{1}H] TROSY spectra of L20C with and without tRNA reveals that the frequencies of most peaks are only marginally affected (Fig 2B). Except for two N-terminal residues (W59 and I60), which disappear, and for D89, the chemical shift changes for all other residues are smaller than 0.07 ppm along the ^{1}H dimension. Although there is a general broadening of the signals which can probably be ascribed to an unspecific electrostatic binding of the small, basic L20C to the large, acidic tRNA (Mr: 26 kDa), no drastic changes are observed in the tRNA/L20C spectrum, whatever the RNA to protein ratio. Altogether, these data support the conclusion that oRNA/L20C and rRNA/L20C form similar, stable and specific complexes, as opposed to the non-cognate unspecific binding of tRNA to L20C.

Analysis of the complex formed with rRNA

Given the large chemical shift perturbation induced in L20C spectrum upon rRNA binding (fig 2C), the direct comparative assignment of the spectrum would have been unreliable. We therefore performed a 3D-NOESY-TROSY experiment on the complex. This experiment allowed us to re-assign about one half of L20C amide cross-peaks (84-88,93,96-100,102-118, full-length L20

numbering). The other 30 spin systems remain too broad to be assigned. They probably correspond to the remaining residues of L20C (59-83, 89-92,94-95). Compared to free L20C, the ^{1}H chemical shift of the amide resonances which could be assigned all fall within 0 to 0.11 ppm from their original value, indicating that these residues are only marginally affected upon RNA binding, whereas most of the broad unassigned resonances corresponded to the strongly shifted peaks. Within the assigned region, the structure of L20C appears largely unchanged: as shown on the strip plot running from Q102 to Q106 for either free or complexed L20C (Fig. 3), the patterns of NOE cross-peaks are very similar, both in terms of frequency and intensity. This indicates that not only the backbone but also the side-chain of these amino acids are unaffected upon rRNA addition. It also confirms the local structure of this region, which remains folded into an α-helix (strong HN_i-HN_{i+1} and HN_{i+3}-$H\alpha_i$ NOEs). Interestingly, these 30 amino acids belong to the C-terminal part of L20C, spanning the two C-terminal α-helices (α3 and α4, light blue, Fig. 4). This is in keeping with observations in the ribosome context, where the surface of interaction between L20C and the 23S rRNA corresponds to the first two helices and loops of the domain. Accordingly, the unaffected C-terminal region of L20C, including α3 and α4, is located on the solvent exposed side of the 50S subunit and is thus not expected to contact the RNA (Fig. 4).
In parallel, the structure of the rRNA within the complex was probed by monitoring the imino proton resonances engaged in base pairs. Both 1D and 2D NOESY experiments were performed on the complex and compared to those recorded with the free rRNA. These indicate that the structure of the rRNA is not globally modified upon L20C binding: the imino/imino, imino/aromatic and imino/amino cross-peaks are nearly identical in the free and the bound rRNA spectra.

The L20C/rRNA complex is larger than expected

Preliminary ^{15}N-^{1}H HSQC experiments performed on the L20C/rRNA and L20C/oRNA complexes revealed that all the amide linewidths were very broad (around 35-40 Hz in the ^{1}H dimension, not shown). This effect was insensitive to the addition of 7 mM $MgSO_4$, which did not significantly modify either the lineshape or chemical shifts of the complex. This is the reason why TROSY-type experiments were used above, as this allowed a partial recovery of the signal. This huge broadening of the L20C peaks upon RNA addition cannot be solely explained by the expected size of this complex (~20 kDa). Other non-exclusive explanations for this increased relaxation could be either conformational or chemical exchange, unspecific aggregation or formation of a higher order molecular structure. In order to investigate this matter, the complex was thus further characterised. The eventual aggregation and/or polydispersity of the NMR sample were investigated by gel-filtration. The RNA/L20C complexes proved to be sufficiently stable to allow purification as a single symmetric peak by size-exclusion chromatography, thereby ruling out significant aggregation or heterogeneity of the formed complex. The presence and integrity of both the RNAs and/or the protein in the purified complex were controlled by gel electrophoresis. NMR experiments performed on the purified complexes (0.15 mM) gave spectra identical to those recorded before purification, with similar very broad lineshapes. This suggested that the complexes corresponded to a higher order molecular structure. This was investigated by small angle X-ray scattering studies (SAXS) on the L20C/rRNA complex. The rationale for focussing on the latter L20C/RNA complex was the following: (i) the structure of L20C complexed to the rRNA is known, at least in the 50S subunit context, and this structural information can be used for an in-depth analysis the SAXS data. (ii) The oRNA 3D structure is unknown; furthermore, it is in slow equilibrium with a minor, alternatively folded, species. This

minor structural heterogeneity of the oRNA does not interfere with the ^{15}N-filtered NMR analyses of the complex, but renders the SAXS data analysis more difficult.

Both the free rRNA and the gel-filtration purified L20C/rRNA complex were analysed on the SAXS beamline D24 at LURE (Orsay, France). The resulting scattering patterns are shown on Figure 5A. The values of the intensity at the origin I(0) and of the radius of gyration R_g were derived from a Guinier analysis (Figure 5A, inset). When compared to that of a BSA calibrating solution the I(0)/c value from the free rRNA curve indicated a molecular mass of the order of 10 kDa, in reasonable agreement with the theoretical Mr of 13 kDa for the rRNA fragment. The scattering curve obtained with the complex indicated that the object present in solution was monodisperse and its I(0)/c value was consistent with the object present in solution being formed of 2 RNA + 2 L20C molecules, *i.e.* a dimer of L20C/rRNA complexes. Higher order multimers of the L20C/rRNA complex would not be compatible with the observed scattering intensity. Complexes with different rRNA/protein ratio could in principle also be formed and give a similar I(0)/c, such as for instance one RNA for three or four proteins. These are however quite unlikely given that, (i) within experimental error, the actual stoichiometry at which the components were mixed was one to one, and, (ii) only one set of protein peaks is seen in the ^{1}H-^{15}N correlation spectra of the complex indicating that there exist only one type of binding site for the protein onto the RNA(s). The radii of gyration were also calculated from the experimental scattering data for both the free RNA and the complex and gave values of 20.4±0.5 Å and 25.7±0.3 Å, respectively. The former value is significantly larger than a theoretical value of 17.8 Å calculated from the rRNA atomic coordinates extracted from the ribosome using the CRYSOL program, suggesting that a small but detectable fraction of RNA is associated in solution as small oligomers. This is confirmed by the appearance of the Guinier plot (see inset to Figure 5A), which exhibits a slight

upward curvature in the smallest angular range (actually the first few points were left out for the Guinier analysis). In contrast, the corresponding plot for the complex shows excellent linearity out to the smallest recorded angle. Finally the pair distribution function of the complex was determined (Figure 5B) yielding a value of the maximum distance of 85±3 Å and a value of the radius of gyration of 26.0±0.1 Å, very close to the previous Guinier estimate. Both values are much larger than the corresponding values of approximately 60 Å and 18.5 Å respectively calculated for a (1:1) complex of L20C and RNA, further supporting the view that the complex dimerises under our conditions.

Preliminary SAXS experiments have also been performed on the L20C/oRNA complex. They yield a value 29.0 ± 0.4 Å for the radius of gyration (Fig. 5A, inset), with a maximum diameter close to 100 Å. These values are slightly larger than the corresponding ones obtained for the L20C/rRNA complex strongly suggesting that the L20C/oRNA complex dimerises in solution like its rRNA counterpart.

Modelling the geometry of the L20C/rRNA dimer

Based on our NMR data, which indicates both that the rRNA fragment is correctly paired and that the surface of interaction of L20C with this fragment is the same as that in the ribosome context, it was reasonable to assume that L20C recognised the rRNA fragment as is observed in the full ribosome context. A complete atomic model of the L20C/rRNA complex was therefore constructed using the 50S crystal structure, as described under experimental procedures. This model was then used to construct a dimer of L20C/rRNA complexes that fitted all the available experimental data. Indeed, based on the results of both the NMR and the SAXS experiments, a number of simple assumptions could be made about the dimer geometry:

(i) *Symmetry constraint:* In the NMR spectra of the L20C/RNA complex, only one single set of peaks is observed for each residue in the protein. This strongly suggests that the dimer is symmetrical, as is usually the case for homodimers, and implies that there must be a two-fold axis relating the two complexes within the dimer.

(ii) *Bridging* constraint*:* L20C or the RNA do not dimerise by themselves. This indicates that the dimer assembly must involve RNA/protein contacts. In other words, within the dimer, one L20C molecule will contact both RNAs (and vice-versa).

(iii) *Exposed C-terminus constraint:* In the TROSY spectrum of the complex, residues within the C-terminal helix of L20C (residues 102-117) appear globally unaffected, in contrast with other parts of the protein, for which large chemical shift displacements are observed. This could be confirmed by tracing the C-terminal helix in the 3D NOESY-TROSY experiment (see above). Thus, the C-terminal helix is unlikely to be in close contact with the other complex in the dimer, but, rather, remains exposed to the solvent.

(iv) *Radius of gyration constraint:* The radius of gyration of the dimer (25.7 Å) is known from the SAXS experiment.

Forming a dimer with two identical monomers related by a two-fold symmetry axis is a problem with 3 degrees of freedom. In order to systematically sample the conformational space, a Monte-Carlo approach was used as described under experimental procedures. The resulting dimers were then selected according to the above geometrical constraints. More specifically, we enforced that the calculated radius of gyration should fall within 24.5 and 27.5 Å, that the bridging distance (see experimental procedures) should be 3 Å or less and that the C-terminus distance (see experimental procedures) should be at least 5 Å. Overall, about 1300 random dimers were generated, 100 of which met the above geometrical criteria. These 100 dimers could be grouped

into only 6 topological families, based on pairwise r.m.s.d. clustering. Each family contained several dimers, suggesting that the sampling process was rather exhaustive. Finally, the ability of the dimers to account for the SAXS data was evaluated. The theoretical scattering profile was calculated for each of the 100 selected dimers and fitted to the experimental data using CRYSOL. Only 8 of the 100 dimers gave high quality fits to the experimental data, with χ values under 2 (see experimental procedures). The best fits to both the scattering profile and the p(r) pair distribution function were obtained with the same model. The corresponding calculated curves are shown superimposed on the experimental data in figure 5A and 5B, respectively. All fits were obtained with a value of the contrast of the hydration shell $\Delta\rho_b = 0.03 \pm 0.003$ electron.Å^{-3}, a very common value corresponding to an average density of about 1.1 within the shell. All models belonged to the same topological family and are shown in figure 6. It thus appears that the above constraints define only one allowable global arrangement for the assembly of L20C/rRNA dimers. The resulting models indeed share a similar topology, in which the two RNA stems form a V, with the helices 40 of the rRNA at the tip, with an angle of 85° ± 15° (Fig. 6). The axes of the two helices 40 are separated by 18 ± 6 Å. The two L20C proteins are located inside the V-shaped assembly. Although these models only provide crude structural information on the possible arrangement within the dimer of complexes, some consistent features are conserved in all of them. For instance, the "bridging" contacts involving one L20C and the rRNA from the other complex always involve residues within loop 2 of L20C (around R90-K91), which are shown as white spheres on figure 6.

Discussion.

The present NMR analysis of the interaction of L20C with RNA fragments corresponding to either its binding site on the 23S ribosomal RNA or to one of its two sites on the operator region of the *rpmI* mRNA shows that this ribosomal protein recognises specifically both nucleic acids in a similar way. The same regions of the protein, helices 1 and 2, are involved in the binding of either RNA, which induce remarkably similar spectroscopic changes. Furthermore, both the oRNA and the rRNA fragments are tightly bound by L20C, as evidenced by the fact the complexes are in slow exchange on the NMR timescale. These results demonstrate that the protein does indeed bind directly its own messenger RNA in a specific manner and provide clear structural evidence for the molecular mimicry model.

More surprisingly, our solution interaction studies show that both L20C/oRNA and L20C/rRNA have very broad NMR lineshapes, which prevented detailed structural analyses. Since both complexes appear long-lived and well defined, as they could be isolated by chromatography, we suspected that the increase in linewidth could result from higher order oligomerisation of the L20C/RNA complexes. To test this possibility, a solution of purified L20C/rRNA complex was studied by small angle X-ray scattering. The object in solution appeared homogeneous and larger than the expected 1 L20C + 1 rRNA complex, both from the value of the intensity at the origin I(0)/c and from the value of the radius of gyration. Indeed, from SAXS data analysis and symmetry considerations, the L20C/rRNA complex most likely dimerises, at least at the concentrations used for either NMR or SAXS studies (~100-200 μM). This finding is quite unexpected, since the assembled ribosome obviously comprises only one L20 protein and one 23S rRNA. More interestingly, a similar behaviour is also observed for the L20C/oRNA complex

where very similar and significant line broadenings are observed in the NMR experiments. The SAXS data further indicates that the L20C/oRNA complex has a slightly larger radius of gyration and maximum diameter than the L20C/rRNA complex, suggesting that it adopts a more extended conformation, *e.g.* the V formed by the two RNA molecules could be more open than that of the L20C/rRNA dimer. All these results suggest that L20C possesses the intrinsic ability to dimerise RNAs that share some resemblance with its ribosomal RNA binding site. Furthermore, this dimerisation ability was observed in stringent conditions, *i.e.* in the absence of divalent cations (Mg^{2+}), which can sometime mediate unspecific aggregation of RNA at high concentrations. This is of great potential relevance in the context of translational control of the IF3 operon for which two target sites for L20 have been identified within the operator region upstream of the *rpmI* translation start, which are both similar to the rRNA binding site of L20C (22). Interestingly, these two sites are able to bind L20 independently, *i.e.* mutations that affect L20 binding at one site do not prevent binding to the other site. However, simultaneous binding at both sites appear to be required for the negative feedback control of *rpmI* (22). This led to the hypothesis that two L20 molecules could be required for the translational repression on the *rpmI* operator. Our observation that L20C is able to promote dimerisation of RNA fragments strongly supports this possibility and provides structural information on the geometry of the resulting complex. The model that was obtained for the dimer suggests that bridging contacts between the two L20C/RNA complexes involve residues within loop 2. Interestingly, loop 2 contains a stretch of residues that are strictly conserved across all bacterial L20 sequences: (D/N)RK. The two basic residues, R90 or K91, would be good candidates as “bridging” residues, as they could contact phosphate groups of the other RNA within the dimer (Figure 6).

Dimer formation was observed using high concentrations of L20C and RNA, which one might argue to be physiologically unrealistic. However, in the context of *rpmI* mRNA, the two L20 binding sites are covalently tethered to one another, thereby increasing their relative apparent concentration. The two L20C binding sites within the operator are separated by approximately 18 nucleotides. Assuming a phosphate to phosphate distance of ~7 Å his means that one site will be constrained to be located within a sphere of *ca.* 120 Å radius centred at the other site. This will result in an apparent concentration of at least 200 µM, *i.e.* similar to that at which dimerisation was observed. This is thus fully consistent with formation of a high order assembly of two L20 molecules on the *rpmI* mRNA, following the model schematically described in Figure 7, which accounts for all the currently available physiological and structural data. The translational repressor would thus be the molecular assembly of two L20C bound to the two sites on the messenger RNA, thereby explaining the simultaneous requirement for two distinct sites. This assembly would be necessary to efficiently hinder the binding of the ribosome to the *rpmI* initiation codon and/or to enhance the stability of the L20/mRNA complex.

Under physiological conditions where L20 concentration becomes limiting, naked 23S rRNA is predicted to displace L20 from its mRNA, thereby de-repressing translation of *rpmI*.

Interestingly, under the above dimer repressor model (Fig. 7), one could also imagine a heterodimer species, in which one rRNA and one mRNA could assemble together with two L20 molecules. In this intermediate complex, in which the rRNA "invades" the dimer structure, the mRNA would become more accessible to ribosomes and thus allow translation to proceed. This could provide a unique and sensitive mechanism to precisely adjust L20 and L35 concentrations to that of the ribosomal RNA. This model predicts that L20 forms a heterodimer assembly with its two binding sites on the mRNA. A direct structural analysis of this complex will require the

"trimming" of the operator region on the extended 5'-leader region of the *rpmI-rplT* mRNA shown in figure 7.

Acknowledgements.

This work was supported by grants from the CNRS "Physique et Chimie du Vivant" and the CNRS/INSERM "Molécules et cibles thérapeutiques" programs, the ACI "Jeunes Chercheurs" of the French Ministry of Research to F.D. S.R. and M.G were recipients of studentships from the French Ministry of Research. The authors wish to acknowledge Dr. Mathias Springer (IBPC, Paris) for support, constant interest and numerous stimulating discussions and Dr. Carine Tisné for the gift of $tRNA^{Lys}_3$.

References

1. Rohl, R., and Nierhaus, K. H. (1982) *Proc. Natl. Acad. Sci. USA.* **79**, 729-733.
2. Herold, M., and Nierhaus, K. H. (1987) *J. Biol. Chem.* **262**, 8826-8833.
3. Yates, J. L., Arfsten, A. E., and Nomura, M. (1980) *Proc Natl Acad Sci U S A* **77**, 1837-1841
4. Yates, J. L., and Nomura, M. (1980) *Cell* **21**, 517-522
5. Zengel, J. M., Mueckl, D., and Lindahl, L. (1980) *Cell* **21**, 523-535
6. Yates, J. L., Dean, D., Strycharz, W. A., and Nomura, M. (1981) *Nature* **294**, 190-192
7. Brot, N., Caldwell, P., and Weissbach, H. (1980) *Proc Natl Acad Sci U S A* **77**, 2592-2595
8. Lesage, P., Truong, H. N., Graffe, M., Dondon, J., and Springer, M. (1990) *J. Mol. Biol.* **213**, 465-475
9. Nomura, M., Gourse, R., and Baughman, G. (1984) *Annu. Rev. Biochem.* **53**, 75-117
10. Lindahl, L., and Zengel, J. M. (1986) *Annu Rev Genet* **20**, 297-326
11. Keener, J., and Nomura, M. (1996) in *Escherichia coli and Salmonella. Cellular and molecular biology.* (Neidhardt, F. C., ed), Second edition Ed., pp. 1417-1431, ASM Press, Washington, DC
12. Said, B., Cole, J. R., and Nomura, M. (1988) *Nucleic Acids Res.* **22**, 10529-10545
13. Gregory, R. J., Cahill, P. B. F., Thurlow, D. L., and Zimmermann, R. A. (1988) *J. Mol. Biol.* **204**, 295-307
14. Philippe, C., Portier, C., Mougel, M., Grunberg-Manago, M., Ebel, J. P., Ehresmann, B., and Ehresmann, C. (1990) *J. Mol. Biol.* **211**, 415-426

15. Serganov, A., Polonskaia, A., Ehresmann, B., Ehresmann, C., and Patel, D. J. (2003) *EMBO J.* **22**, 1898-1908

16. Scott, L. G., and Williamson, J. R. (2001) *J. Mol. Biol.* **314**, 413-422

17. Serganov, A., Ennifar, E., Portier, C., Ehresmann, B., and Ehresmann, C. (2002) *J. Mol. Biol.* **320**, 963-978

18. Nikulin, A., Serganov, A., Ennifar, E., Tishchenko, S., Nevskaya, N., Shepard, W., Portier, C., Garber, M., Ehresmann, B., Ehresmann, C., Nikonov, S., and Dumas, P. (2000) *Nat. Struct. Biol.* **7**, 273-277

19. Agalarov, S. C., Sridhar Prasad, G., Funke, P. M., Stout, C. D., and Williamson, J. R. (2000) *Science* **288**, 107-113

20. Lesage, P., Chiaruttini, C., Graffe, M., Dondon, J., Milet, M., and Springer, M. (1992) *J. Mol. Biol.* **228**, 366-386

21. Chiaruttini, C., Milet, M., and Springer, M. (1996) *EMBO J.* **15**, 4402-4413

22. Guillier, M., Allemand, F., Raibaud, S., Dardel, F., Springer, M., and Chiaruttini, C. (2002) *RNA* **8**, 878-889.

23. Raibaud, S., Lebars, I., Guillier, M., Chiaruttini, C., Bontems, F., Rak, A., Garber, M., Allemand, F., Springer, M., and Dardel, F. (2002) *J Mol Biol* **323**, 143-151.

24. Harms, J., Schluenzen, F., Zarivach, R., Bashan, A., Gat, S., Agmon, I., Bartels, H., Franceschi, F., and Yonath, A. (2001) *Cell* **107**, 679-688.

25. Noller, H. F., Kop, J., Wheaton, V., Brosius, J., Gutell, R. R., Kopylov, A. M., Dohme, F., Herr, W., Stahl, D. A., Gupta, R., and Woese, C. R. (1981) *Nucleic Acids Res.* **9**, 6167-6189.

26. Tisne, C., Rigourd, M., Marquet, R., Ehresmann, C., and Dardel, F. (2000) *RNA* **6**, 1403-1412

27. Pervushin, K., Riek, R., Wider, G., and Wüthrich, K. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 12366-12371

28. Zhu, G., Kong, X. M., and Sze, K. H. (1999) *J. Biomol. NMR* **13**, 77-81

29. Rouh, A., Delsuc, M. A., Bertran, G., and Lallemand, J. Y. (1993) *J. Mag. Reson.* **102**, 357-359

30. Boulin, C., Kempf, R., Koch, M. H. J., and McLaughlin, S. M. (1986) *Nucl. Instrum. and Meth.* **A249**, 399-407

31. Dubuisson, J. M., Decamps, T., and Vachette, P. (1997) *J. Appl. Cryst.* **30**, 49 - 54.

32. Guinier, A., and Fournet, G. (1955) *Small Angle Scattering of X-Rays*, Wiley, New York

33. Svergun, D. I. (1992) *J. Appl. Crystallogr.* **25**, 495-503

34. Svergun, D. I., Barberato, C., and Koch, M. H. J. (1995) *J. Appl. Crystallogr.* **28**, 768-773

35. Brünger, A. T. (1992) *X-PLOR Version 3.1. A system for X-Ray Crystallography and NMR.*, Yale University Press, New Haven CT.

36. Brooks, B., Bruccoleri, R., Olafson, B., States, D., Swaminathan, S., and Karplus, M. (1983) *J. Comp. Chem.* **4**, 187-217

37. Zengel, J. M., Vorozheikina, D., Li, X., and Lindahl, L. (1995) *Biochem Cell Biol* **73**, 1105-1112

Figure Legends.

Figure 1: Sequences of the RNA fragments used in binding studies.

rRNA:L20C binding site on *E. coli* 23S rRNA, as deduced from the structure of *D. radiodurans* 50S ribosomal subunit (24), numbering is that of the *E. coli* 23S rRNA. oRNA: Secondary structure of the L20 binding site 2 on the *rpmI* translational operator(22). Outlined residues were added to increase the pairing stability.

Figure 2: ^{1}H-^{15}N spectra of free and RNA-bound L20C.

(A) HSQC spectrum of free L20C. Peaks marked with a star correspond to spin systems that were not assigned in the full-length L20. (B) (C) and (D) : Superimposition of the TROSY spectra of L20C/RNA equimolar mixtures (red) onto the HSQC spectrum free L20C (black) : (B) : L20C/$tRNA^{Lys}_{3}$. (C) L20C/rRNA complex. (D) L20C/oRNA complex. Similarly shifted peaks in the L20C/oRNA and L20C/rRNA complexes are boxed.

Figure 3: NMR Assignment of the L20C C-terminal helix in the free and RNA-bound forms. Strip-plots were taken from ^{15}N-NOESY-HMQC (A) and ^{15}N-NOESY-TROSY (B) experiments of the free and rRNA-bound L20C, respectively. Shown are the spin-systems corresponding to region Q102-Q106. Sequential connectivities are indicated with the corresponding cross-peaks shown in green, while internal cross-peaks are shown in red.

Figure 4: Chemical shift perturbation mapping of RNA binding on L20C.

Shown is the schematic structure of the complex L20C/rRNA complex, extracted from the crystal structure of the 50S subunit(24). The chemical shifts of residues shown in cyan are either mildly or not affected by either rRNA or oRNA binding, whereas peaks corresponding to residues

shown in red are either strongly affected or broadened down to baseline. The N- and C-terminus of L20C are indicated.

Figure 5: SAXS analysis of the L20C/RNA complexes.

Panel A: Scattering patterns from the free RNA (open circles) and the L20C/rRNA complex (squares) The calculated scattering pattern for the best-fitting model of the complex is shown as a dashed line. Inset : Guinier plots for the free RNA (open circles), the L20C/rRNA complex (squares) and the L20C/oRNA complex (filled circles). Panel B: Pair distribution function p(r) for the L20C/rRNA complex. Dots : experimental curve, solid line : p(r) calculated from the best-fitting model.

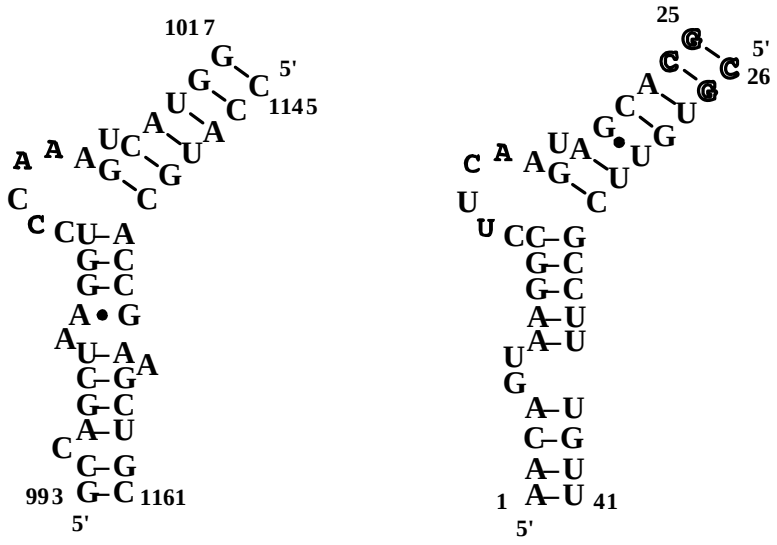
Figure 6: Model of the L20C/rRNA dimer structure.

The space of allowable geometries for L20C/rRNA dimers was sampled as described under experimental procedures. Only those that fitted the experimental SAXS data were retained. Shown on the left are the superimposed backbones of the 8 resulting RNA/L20C dimers. On the right is shown a ribbon drawing of the best-fitting structure. Bottom view and top view are rotated by 90 degrees about a horizontal axis. One L20C/rRNA complex is shown in yellow/cyan, and the other in red/green. Potential bridging residues, located in loop 2, are shown as white spheres.

Figure 7: model of *rpmI-rplT* translational control.

Left view: schematic drawing of 2D structure of the 5' leader region of the *rpmI-rplT* operon, that has been determined previously(21). The stop codon of *infC* and the ribosome binding site and start codon of *rpmI* are boxed. The two L20 binding sites(22) are shown in green. In this representation, the two halves of the pseudoknot are not paired.

Right view: Folded view of the operator, after a 180 degrees rotation of the right part of the structure about the dashed axis on the left. Two molecules of L20 (blue) would bind simultaneously to the two sites and form a large molecular assembly sequestering *rpmI* translation signals



rRNA oRNA

Figure 1

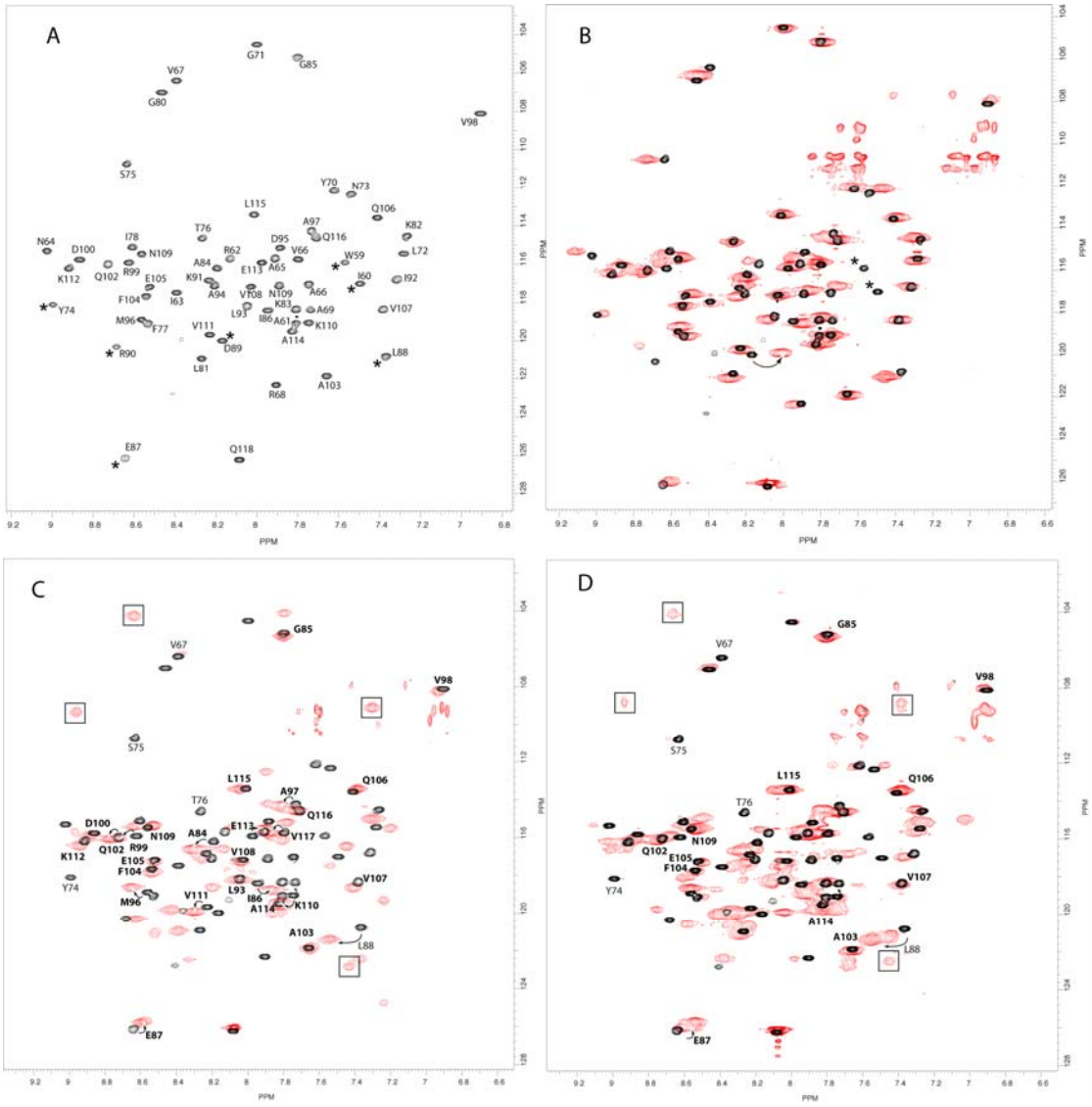


Figure 2

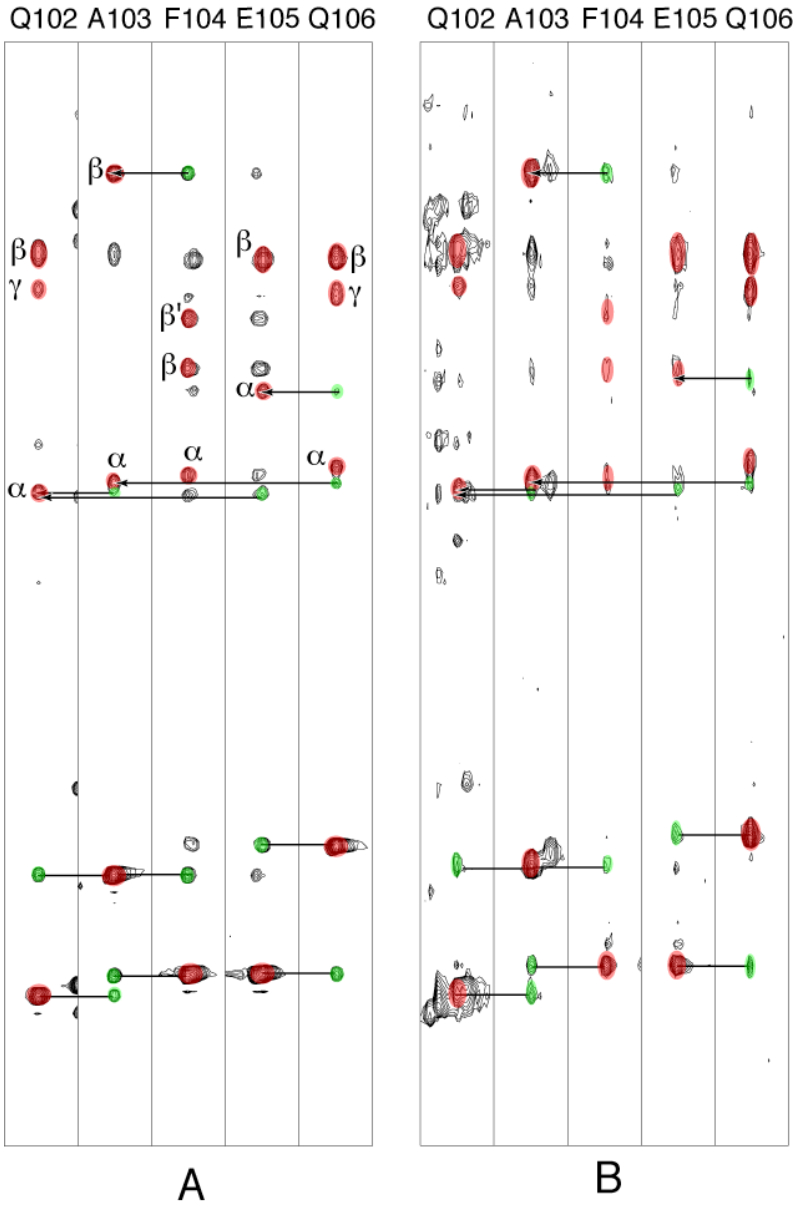

Q102 A103 F104 E105 Q106
β
β
γ
β'
β
α
α
α
α
A
Q102 A103 F104 E105 Q106
B

Figure 3

Figure 4

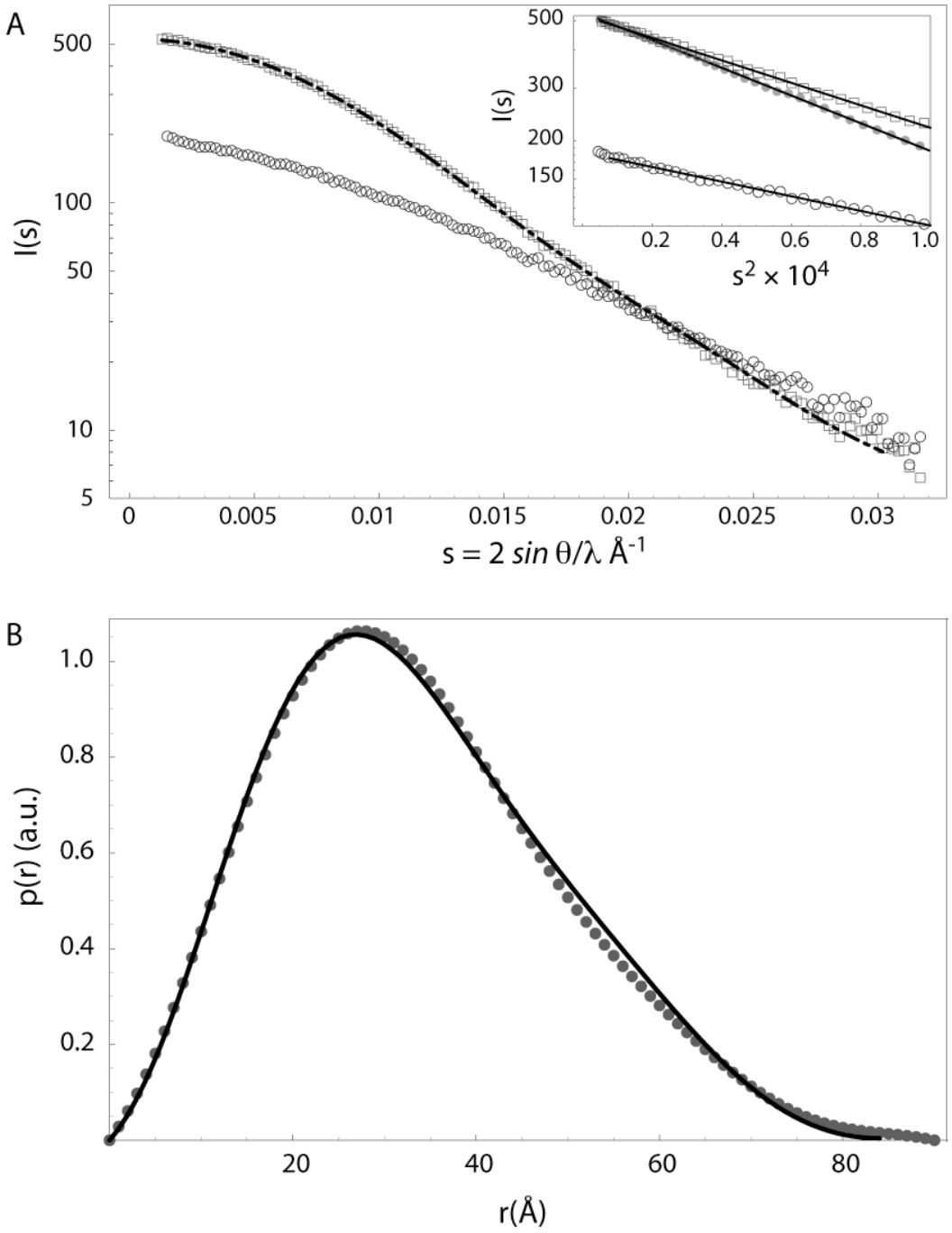

A
500
100
50
10
5
I(s)
0
0.005
0.01
0.015
0.02
0.025
0.03
s = 2 sin θ/λ Å-1
500
300
200
150
I(s)
0.2
0.4
0.6
0.8
1.0
s2 × 104
B
1.0
0.8
0.6
0.4
0.2
p(r) (a.u.)
20
40
60
80
r(Å)

Figure 5

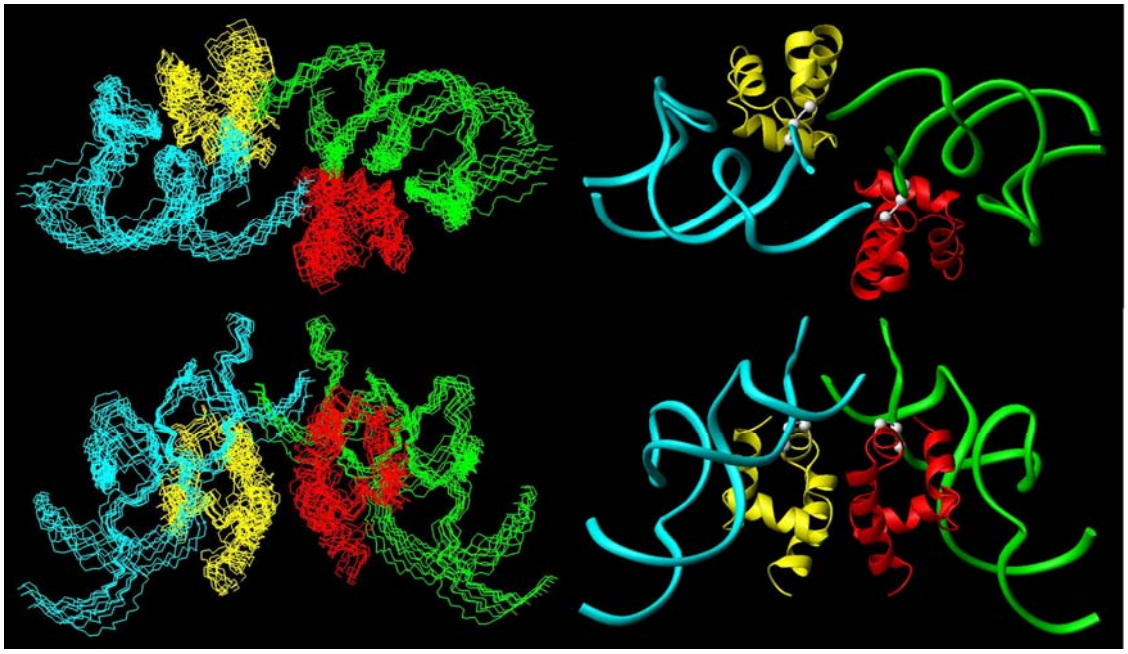


Figure 6

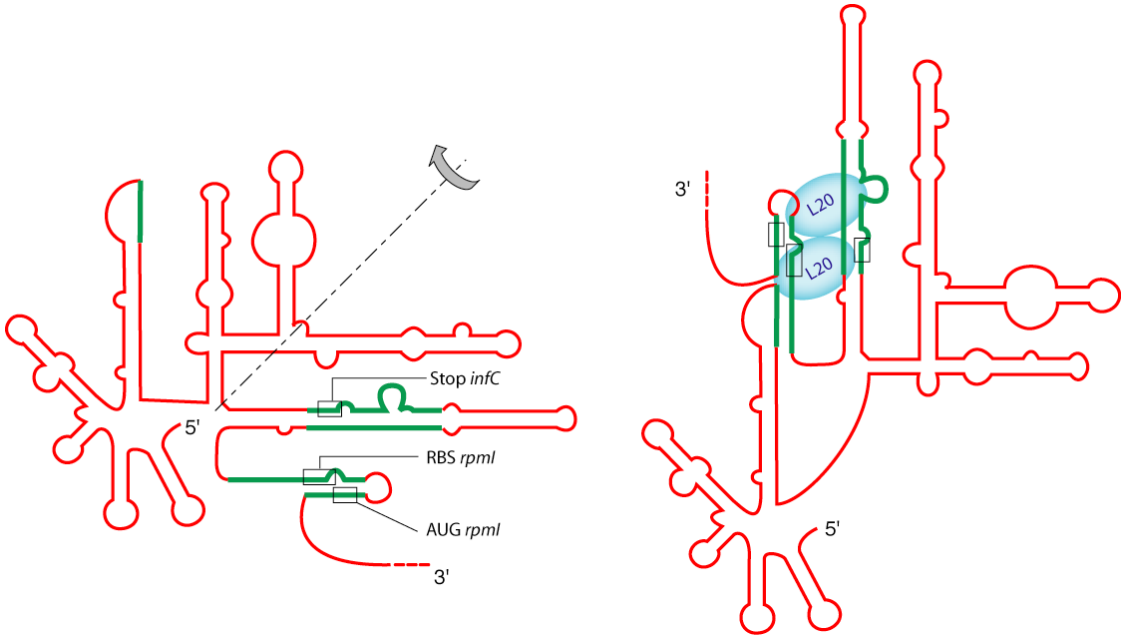


Figure 7